
知识管理论坛
2022 年 第 1 期(总第 37 期)
ChinaXiv合作期刊

【学术探索】

近 20 年计算机与信息科学领域研究进展
——*IPM* 期刊主题分析

李涵霄[1,2] 杜杏叶[1,2]
1. 中国科学院文献情报中心 北京 100190
2. 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190

摘要：[目的 / 意义] 对《信息处理与管理》2000-2020 年刊载论文进行主题分析，以了解 *IPM* 期刊的主题侧重与演进趋势，为计算机与信息科学领域发展及相关研究提供参考。[方法 / 过程] 首先，基于 ScienceDirect 全文数据库中的 1 852 篇研究论文，对论文标题、摘要与关键词进行统计与可视化，划分主题大类；其次，对各类别的研究主题进行系统梳理；最后，比较不同时期的研究重点，分析主题演进趋势。[结果 / 结论] *IPM* 主要关注信息检索、文本分析、用户研究 3 类主题，总体呈现出始终以信息检索为核心主题、从文本与信息分析向多媒体与知识分析转变、对用户情感的深入分析与挖掘等演变特征。

关键词：《信息处理与管理》 *IPM* 计算机与信息科学 主题分析
分类号：G201

引用格式：李涵霄 , 杜杏叶 . 近 20 年计算机与信息科学领域研究进展 : *IPM* 期刊主题分析 [J/OL]. 知识管理论坛 , 2022, 7(2): 24-36[引用日期]. http://www.kmf.ac.cn/p/272/.

1 引言

数字化、网络化与智能化时代的到来，为计算机与信息科学领域的发展带来了巨大冲击，也使得相关研究迅速增多，研究主题发生变化。作为科研成果的主要载体和传播平台，科技期刊在学术交流中承担着重要的使命[1]。对科技期刊载文进行主题分析，能够更好地了解该学科领域的研究进展与演进特征。《信息处理与管理》（*Information Processing and Management*，*IPM*）于 1963 年创刊，其最初名为《信息存储与检索》（*Information Storage and Retrieval*，*ISR*），自 1975 年正式更名为 *IPM*，并延续至今。根据 SCI-JCR 数据，该刊 2020 年的引用评分（CiteScore）为 8.6，影响因子为 6.222，在计算机科学与信息系统（COMPUTER SCIENCE，INFORMATION SYSTEMS）以及情报学与图书馆学（INFORMATION SCIENCE & LIBRARY SCIENCE）类别中均位列一区。*IPM* 致力于发

基金项目：本文系中国科学院自然科学期刊编辑研究会 2020 年资助项目“中国期刊画像及建设世界一流期刊发展策略研究”（项目编号：YJH003）研究成果之一。
作者简介：李涵霄，硕士研究生；杜杏叶，副研究馆员，副编审，博士，硕士生导师，通信作者，E-mail: duxy@mail.las.ac.cn。

收稿日期：2021-10-12 发表日期：2022-02-16 本文责任编辑：刘远颖

知识管理论坛，2022（1）:24-36
DOI: 10.13266/j.issn.2095-5472.2022.003

表计算机与信息科学交叉领域的前沿研究成果，在国内外计算机与信息系统界乃至图书情报界均具有高影响力与高知名度，为推动领域进步做出了重要贡献。因此，对 *IPM* 期刊的发文主题进行系统梳理，能够在一定程度上反映计算机与信息科学领域的研究进展，展现 *IPM* 为领域发展所做出的学术贡献。

已有学者针对 *IPM* 进行了主题分析。F. E. DeHart 对 *IPM*、*JASIS*（*Journal of the American Society for Information Science*）和 *JD*（*Journal of Documentation*）3 个期刊 1987-1990 年所发表论文的参考文献进行分析，着重比较了引用专著的比例，发现 *IPM* 在 1989-1990 年引用专著最多的 3 个主题分别为信息存储和检索系统、人工智能、话语分析 [2]；M. Y. Tsay 对 1998-2008 年 *JASIST*（*Journal of the American Society for Information Science and Technology*）、*IPM* 和 *JD* 3 个期刊进行文献计量分析与比较，发现 *IPM* 引用期刊论文最多的 3 个主题分别为搜索、在线信息检索、信息工作，引用书籍最多的 3 个主题分别为信息存储和检索系统、信息检索、计算机算法 [3]；王曰芬等对 2006-2015 年《现代图书情报技术》及 *IPM* 等国内外相似期刊的发文特征进行比较分析，发现信息检索是 *IPM* 期刊最大的研究热点，其他热点还有用户行为分析、文本挖掘算法、文本分类、语义分析等 [4]。可以看出，此前研究均是将 *IPM* 与其他期刊进行比较分析，且大多是定量分析，而较少关注各个研究主题的内涵演变。因此，笔者对 *IPM* 近 20 年（2000-2020 年）的发文主题进行系统梳理，以了解 *IPM* 期刊发文的主题侧重与演进趋势，为计算机与信息科学领域发展及相关研究提供参考，也为图书情报领域提供有益借鉴。

2 数据与方法

2000-2020 年间，*IPM* 共刊发 1 852 篇研究论文。利用 ScienceDirect 全文数据库将论文数据导出，形成可供统计分析的数据源。首先，对论文关键词词频进行统计，绘制关键词词云图（见图 1），发现研究热点主要涉及信息检索（information retrieval）、机器学习（machine learning）、自然语言处理（natural language processing）、查询扩展（query expansion）、社交媒体（social media）、情感分析（sentiment analysis）、文本分类（text classification）、信息搜寻（information seeking）等主题。其次，提取论文标题与摘要，利用 VOSviewer 进行共现分析，得出共现网络图（见图 2），从图 2 中可以明显看出，研究主题主要分为三大类：信息检索（information retrieval）、文本分析（text analysis）和用户研究（user research）。最后，基于 1 852 篇研究论文，从信息检索、文本分析和用户研究三大类别出发，对 *IPM* 近 20 年的研究主题进行梳理，并比较不同时期的研究重点，得出主题演进趋势。



图 1 2000-2020 年 *IPM* 关键词词云

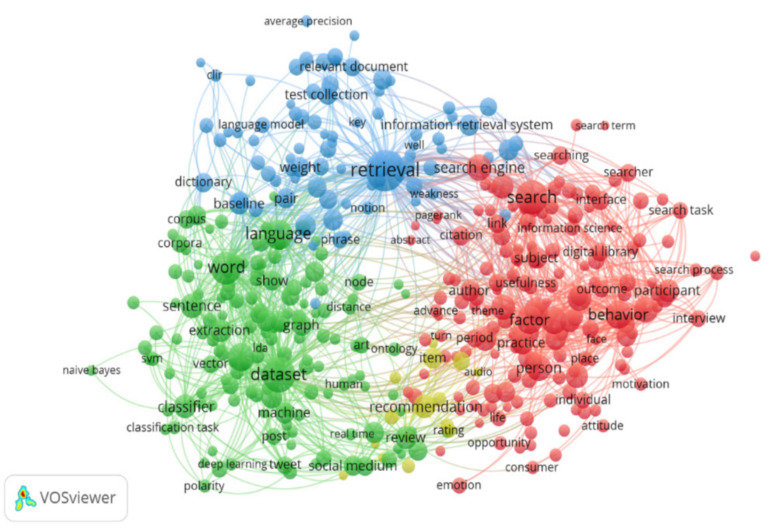


图 2 2000-2020 年 *IPM* 标题与摘要共现网络

3 *IPM* 主题分析

3.1 第一大主题：信息检索

信息检索是 20 世纪 50 年代在国外兴起的一门新兴学科，主要研究的是信息的表示、存储、组织与访问 [5]。在过去 20 年中，信息检索一直是 *IPM* 期刊关注的重点，涉及的主题包括信息检索模型（information retrieval model）、搜索引擎（search engine）、图像检索（image retrieval）等方面。

3.1.1 信息检索模型

信息检索模型指描述信息检索中的文档、查询和它们之间关系（匹配函数）的数学模型 [6]，常用的检索模型有布尔检索模型、概率模型、向量空间模型、语言模型、排序模型等类型。*IPM* 相关研究基本围绕概率模型展开，如 K. Sparck Jones 等开发了信息检索概率模型 [7-8]，该模型是对其团队于 1976 年提出的概率模型的改进 [9]，也是目前最广受认可的检索模型，许多研究在该模型的基础上进行改动，形成了较为普遍的应用形式，见公式（1）[10]：

$$sim(Q,D)=\sum_{q\in Q}\log\frac{(r_i+0.5)/(R-r_i+0.5)}{(n_i-r_i+0.5)/(N-n_i-R+r_i+0.5)}\cdot\frac{(k_1+1)f_i}{K+f_i}\cdot\frac{(k_2+1)qf_i}{k_2+qf_i}\qquad \text{公式 (1)}$$

其中，f_i 指检索文档中词项 i 出现的次数，qf_i 指查询中词项 i 出现的次数，N 指整个检索文档数据集合的大小，r_i 指包含词项 i 的相关文档的数量，n_i 指包含词项 i 的文档的数量，R 指和查询相关的文档集合的大小，K、k_1、k_2 均为根据经验设定的超参数 [10]。

其他学者也从多个角度出发构建了信息检索概率模型，并进行了实验测试。Z. B. Xu 等开发了一种基于狄利克雷复合多项式（Dirichlet Compound Multinomial，DCM）分布的信息检索概率模型，能够实现高效检索和准确排名 [11]；F. Dahak 等针对 XML 信息检索，建立了一个利用用户期望来估计上下文重要性的概率模型，并通过实验证明了其有效性 [12]。此外，向量空间模型与排序模型也受到了广泛讨论，如 X. Y. Tai 等提出一种基于向量空间模型（VSM）的信息检索模型，能够利用用户提供的相关信息来提高检索性能 [13]；J. F. Guo 等分析比较了神经排序模型的基本假设、设计原则和学习策略，并讨论了学习索引、利用外部知识学习、利用可视化技术学习、利用语境学习、神经排序模型理解等未来发展趋势 [14]。整体而言，相关研究主要关注检索模型的构建、测试、比较与改进，目的是提升信息检索系统性能，实现更高精度的信息检索。

DOI: 10.13266/j.issn.2095-5472.2022.003

3.1.2 搜索引擎

在搜索引擎方面，相关研究集中在搜索引擎的检索性能评估、查询扩展和相关反馈以及网页排名算法等方面。关于检索性能评估，L. Vaughan 基于查准率和查全率的概念，提出一组测量方法，用于评估搜索引擎性能与稳定性，实验结果表明该测量方法能够有效区分搜索引擎性能[15]；F. Can 等认为人为评估搜索引擎检索性能成本较高，因此引入一种自动搜索引擎评估方法，并通过实验证明其评估结果与人为评估一致[16]。作为查询优化的重要分支，查询扩展主要通过用户查询日志与用户相关反馈等来源中的信息，对用户的查询进行扩展，如 H. Kim 等提出了一种基于用户查询日志聚类的方法，能够在一定程度上弥补用户查询和检索系统间的词汇鸿沟[17]；S. Jung 等以搜索引擎用户的点击数据作为隐式相关反馈的信息来源，讨论了相关反馈的可靠性及其变化[18]。此外，查询扩展还更加关注基于语义的相关反馈技术，以应对查询和文档间的语义鸿沟，如 J. M. Wang 等提出了一种结合相关匹配和语义匹配的伪相关反馈，以提高反馈文档质量[19]。在网页排名算法方面，在 PageRank、HITS 和 OPIC 等主流算法的基础上，众多学者开发了更高效的排名算法，如 A. M. Z. Bidoki 等提出了一种基于强化学习的 DistanceRank 算法，将两个网页间的“平均点击次数”定义为距离，距离较小的页面能够具有更高的排名，实验结果表明该算法在网页排名和抓取调度方面优于其他算法[20]。

3.1.3 图像检索

在文本检索的基础上，以图像、音频、视频作为检索对象的多媒体检索技术逐渐发展起来，*IPM* 相关研究则主要集中在图像检索领域。2000 年后，图像检索从基于文本的图像检索（Text-based Image Retrieval，TBIR）向基于内容的图像检索（Content-based Image Retrieval，CBIR）发展。CBIR 的基础即是对图像的颜色、纹理、形状等内容特征进行选择、提取和表示[21]，相关研究也基本从该角度出发。如 P. W. Huang 等基于纹理相似度，提出了两种纹理特征表示方法（CSG-vector 和 EDP-string），并据此设计了高效的图像检索系统[22]；T. C. Lu 等针对图像的颜色特征，以颜色分布、平均值和标准差表示图像的全局特征，以图像位图表示图像的局部特征，以提高图像检索的准确性[23]。然而，基于内容的图像检索也存在难以跨越的语义鸿沟[24]，为此，基于语义的图像检索技术逐渐发展起来，S. Pandey 等便提出了一种用于语义分类分层图像数据库的语义和图像检索系统，使得图像被映射到多维特征空间的同时，图像语义也能够通过聚类和索引被表示出来，最终实现所需语义和对应图像的高效检索[25]。

3.2 第二大主题：文本分析

文本分析即对文本内容进行表示和特征提取，使得文本能够被计算机识别与处理，从而判断文本主题以及文本提供者的态度和情绪。*IPM* 中有关文本分析的研究主要集中在文本挖掘（text mining）、情感分析（sentiment analysis）、知识图谱（knowledge graph）等方面。

3.2.1 文本挖掘

针对潜藏于电子形式中的大量文本数据，文本挖掘能够从中抽取事先未知的、可理解的、最终可用的知识，并运用这些知识更好地组织信息以支持参考利用[26-27]。*IPM* 相关研究基本围绕文本分类与文本聚类展开。其中，文本分类指将文档组织为预先定义好的类别，通常使用机器学习算法[28]，如 A. Elnagar 等比较了常用的阿拉伯语文本分类深度学习模型，并提出了一个完全基于深度学习模型的分类方法[29]。同时，文本分类针对的文本特征也从简单的词、短语和句子发展为语法和语义特征，如 A. Mohasseb 等针对问答系统中的问题分类，提出了一种基于语法的分类框架，能够有效区分不同的问题类型[30]；Z. Kastrati 等提出了一种语义丰富的文档表示模型，能够对金融文档进行自动分类[31]。在文本聚类方面，众多学者提出了各种聚类算法以优化聚类性能，如 G. B. Hu 等开发了一种基于 K-Means 聚类算法的半监督

聚类方法，能够对聚类过程进行约束[32]；C. L. Chen 等提出了一种基于频繁模糊项集的分层聚类方法，旨在提高分层聚类精度[33]；还有学者提出了用于文档聚类的概率模型与算法，并通过实验证明其性能优于此前广泛使用的模型与算法[34-35]。此外，也有研究从应用场景出发，探讨了文本挖掘在信息检索[36]、用户服务[37]、专利分析[38]、话题识别[39]等领域的应用方法与实践效果。

3.2.2 情感分析

论坛、博客等各类社交媒体的发展以及以大众点评为代表的点评网站的出现，为大众提供了情绪交流与消费点评的开放式平台[40]，也因此产生了大量的针对产品、服务、事件、话题等实体的观点、情感、评价、态度与情绪[41]。情感分析，或称观点挖掘，便是利用自然语言处理和文本挖掘技术，对这些带有情感色彩的主观性文本进行分析、处理和抽取的过程[42]。*IPM* 中的情感分析研究主要以社交媒体为依托平台，从用户发布内容或评论中分析其观点和情绪，如 A. Balahur 等与 S. M. Mohammad 等以 Twitter 为例，分析了推文中的情感、情绪、目的、风格以及相应的情感分析系统[43-44]；A. Severyn 等针对 YouTube 上大量的用户生成内容，构建了能够应对新领域或新语言的观点挖掘模型，并通过实验进行了验证[45]。还有学者分析了用户对产品或服务的评价，以挖掘其中的态度与情绪，如 M. Al-Smadi 等提出了一种基于监督机器学习的方法，能够对酒店评论进行情绪分析[46]。在情感分析的过程中，研究者构建了许多用于不同场景的情感分析模型，如 A. Kumar 等提出了一种用于文本和视觉社交数据中细粒度情感分析的深度学习模型[47]；Z. Mahmood 等则开发了罗马乌尔都语语料库，并以此为依据开发了一种用于挖掘情绪和态度的深度学习模型[48]。

3.2.3 知识图谱

知识图谱本质上是揭示实体 / 概念之间语义关系的语义网络[49]。*IPM* 相关研究主要涉及知识图谱技术、知识图谱构建与知识图谱应用 3 个方面。在知识图谱技术方面，众多学者以知识实体的抽取为主，探讨了知识图谱中的知识抽取与知识表示，如 H. C. Cho 等研究了多段表示的命名实体识别[50]；L. Derczynski 等描述了一个 Twitter 实体消歧数据集，并对推文中的命名实体识别和消歧进行了实证分析[51]；X. Tang 等提出了一种多源知识表示学习的模型，以结合实体描述、层次类型和文本关系，提高知识表示有效性[52]。在知识图谱构建方面，相关研究主要以语料库为基础，构建基于语义关系的知识图谱，如 I. Bounhas 等从阿拉伯语的有声语料库中构建了一个形态语义知识图谱，利用上下文知识来推断实体之间的语义依赖关系，并评估了文档索引和查询扩展的集中使用场景[53]。在应用方面，知识图谱可以应用至检索系统、问答系统、大数据分析等领域，如 D. F. Li 等提出了一种级联模型，能够同时考虑语义特征和图谱特征，并设计了不同的级联结构，以用于知识推理和检索[54]；S. Shin 等针对问答系统中咨询问题的含义，设计了一种谓词约束词典，并提出了基于该种谓词约束的问答系统，能够提高搜索准确性[55]；F. Janssens 等选择了图书情报领域的 5 本期刊，对其 2002-2004 年间刊载的近千篇文献进行了计量分析，利用知识图谱绘制了可视化术语网络[56]。

3.3 第三大主题：用户研究

对用户的信息获取、查寻、利用等行为进行研究，有助于信息服务机构更具针对性地改进信息服务系统性能，提升服务质量[57]。信息搜寻行为（information seeking behavior）、用户生成内容（user generated content）、个性化服务与人机交互（personalized service & human-computer interaction）是用户研究主要关注的主题。

3.3.1 信息搜寻行为

信息搜寻行为指个体为满足某些目标需求而有目的地搜寻信息，除包括普遍意义上的信息搜索外，信息搜寻更侧重于满足整个信息需求的完整过程，探索用户搜寻行为背后的原因、

知识管理论坛，2022（1）:24-36
DOI: 10.13266/j.issn.2095-5472.2022.003

影响因素、用户特征和个人差异等方面[58]。许多学者针对不同类型的用户，分析了其信息搜寻行为的特征，如 S. Makri 等通过对 27 位律师的信息搜寻行为进行分析，提出了对 Ellis 信息搜寻行为模型的改进[59]；H. R. Jamali 等对物理和天文学研究人员的信息搜寻行为进行了调查，揭示了不同学科在信息搜寻行为上的差异，并发现跨学科领域更有可能使用通用搜索工具来获取信息[60]；M. Lykke 等调查了医生的信息搜寻行为，发生多数医生能够利用系统特征和搜索策略生成结构良好的查询[61]。此外，随着社会水平的不断提高，人们对健康信息的需求也逐渐增多，健康信息行为受到了更加广泛关注，相关研究从不同角度着手分析了健康信息行为。如针对健康信息需求，W. J. Pian 等系统梳理了消费者健康信息需求理论，指出未来应关注的社会和情感维度[62]；针对健康信息提供方，X. F. Zhang 等探讨了医生在网络平台分享健康信息的动机，发现除物质动机外，职业动机起着主要作用[63]；针对健康信息获取方，I. Huvila 等研究了中老年人的健康信息行为偏好和动机，并将其与年轻人和老年人的健康信息行为进行了比较[64]。还有研究对健康信息规避行为进行了分析[65]，以促进人们对健康信息的搜寻。

3.3.2 用户生成内容

用户生成内容是随着 Web 2.0 兴起而发展起来的一种网络信息资源创作与组织模式，指用户以各种形式在网络上创作的文字、图片、视频等内容[66]。*IPM* 相关研究较少直接讨论用户生成内容的理论基础，而是将其作为观点挖掘、情感分析、舆情管理等领域研究的数据来源来进行分析。如 A. Severyn 等对 YouTube 上的用户生成内容进行了观点挖掘[45]；Y. D. Ge 等探讨了用户生成内容中的情绪对股市的影响[67]；L. F. Li 等研究了自然灾害发生后社交媒体上公众的负面情绪，以及具有大量追随者的用户的发言对转发数量的影响[68]。此外，社交媒体上的用户生成内容还可能会造成谣言和虚假新闻的传播，学者们对谣言的识别和检测进行了研究。Y. H. Liu 等提出了一种基于长短期记忆网络（Long Short-Term Memory，LSTM）和最大池化（max pooling）的模型，通过捕获转发内容、传播者和传播结构的动态变化来识别谣言传播过程，并利用新浪微博数据进行了验证[69]；S. A. Alkhodair 等则提出了一种基于 word2vec 和长短期记忆循环神经网络（LSTM-RNN）的突发性谣言检测模型（见图 3），并利用 Twitter 中的数据进行了实验，证明该模型在查准率、查全率及 F1 值等方面的性能优于其他模型[70]。

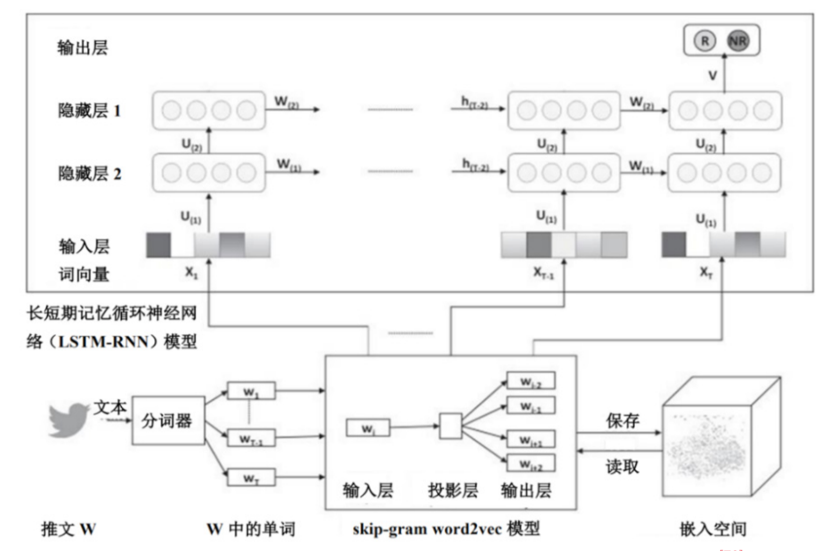


图 3 基于 word2vec 和 LSTM-RNN 的突发性谣言检测模型[70]

在该模型中，首先，推文 W 被标记为一系列单词（W_1、⋯、W_{T-1}、W_T）；然后，word2vec 模型将单词序列转换为向量序列（X_1、⋯、X_{T-1}、X_T），并通过加权连接将其传递给 LSTM-RNN 模型；最后，LSTM-RNN 模型将判断样本类别为谣言（R）或非谣言（NR），并将其作为最终的输出向量 [70]。

3.3.3 个性化服务与人机交互

在个性化服务方面，相关研究主要围绕推荐系统的个性化展开。J. Wang 等提出了社交媒体个性化框架，并构建了一个排名模型，能够在标签和内容推荐中集成用户的标记历史，使得系统建议与用户偏好保持一致 [71]；F. M. Belem 等以对象和用户为中心，对个性化标签推荐进行了改进，并将其与以对象为中心的推荐方法进行了比较 [72]；S. Renjith 等讨论了旅游推荐系统的发展，梳理了从通用搜索引擎到个性化推荐系统再到基于情境感知的个性化推荐系统的演变 [73]。在人机交互方面，除讨论用户和互联网资源 [74]、信息检索系统 [75] 的交互，以及人机交互中的情感 [76] 等方面外，眼动跟踪技术也是广受关注的主题。如 M. J. Cole 等利用眼动跟踪技术，对用户的交互式信息获取过程进行了建模，以预测用户的知识水平 [77]；M. Clark 等基于眼动跟踪数据，分析了用户对电子邮件文本的交互方式 [78]；B. Hilberink-Schulpen 等通过眼动跟踪方法调查了招聘广告中外语的使用是否会影响用户的注意力和观看方式 [79]。

4 *IPM* 主题演进趋势

在不同的时间段内，*IPM* 的研究主题有着不同的侧重。为了解不同时期的研究重点，并从整体上分析论文主题的演进趋势，笔者将 2000-2020 年的论文划分为 4 个时间段，对每个时间段内的关键词词频进行统计与可视化，制作不同时期的词云图（见图 4）。在 4 个时间段内，信息检索（information retrieval）始终是最主要的研究主题，这与 *IPM* 期刊的定位密切相关。然而，随着时间的推移，在 2016-2020 年间，信息检索的研究热度下降，社交媒体（social media）、情感分析（sentiment analysis）、深度学习（deep learning）、机器学习（machine learning）、自然语言处理（natural language processing）、文本挖掘（text mining）等主题词的频率增高，其中社交媒体以 33 次的词频，超过了信息检索（31 次），成为研究热度最高的主题词。



图 4 2000-2020 年不同时期 *IPM* 关键词词云

ChinaXiv合作期刊
知识管理论坛，2022（1）:24-36
DOI: 10.13266/j.issn.2095-5472.2022.003

利用 CiteSpace 软件进行突现词探测（burst detection），以展现研究主题的发展脉络和演进趋势，并预测未来的研究方向。2000-2020 年 *IPM* 突现词、突现强度及起止年份见表 1，从表 1 中可以看出，在近 20 年的时间中，共有 24 个突现词，其中“信息检索”（information retrieval）的研究热度从 2000 年开始，持续到 2009 年；“排名”（ranking）的研究热度从 2012 年开始，持续到 2020 年。其他持续时间较长（5 年以上）的突现词还有：万维网（world wide Web）、检索（retrieval）、相关性（relevance）、设计（design）、共引（cocitation）、用户（user）等。目前的研究热点及前沿包括: 排名（ranking）、图像（graph）、影响（impact）、知识（knowledge）、Twitter、情感分析（sentiment analysis）、口头语言（word of mouth）、情绪（emotion）等。这与此前的分析结果大致吻合，即信息检索在很长时间内一直是 *IPM* 主要关注的内容，同时由于搜索引擎的发展离不开网页排名算法的更新与优化，因此排名算法在持续较长时间的研究热度后，仍然是目前的研究前沿。此外，研究主题的演变还呈现出从文本到图像、从信息到知识的特征。从表现形式来看，信息检索的对象逐渐脱离文本的限制，而开始关注具有更丰富信息的图像、音频与视频，这些多媒体载体除包含基本的文本信息外，还能够传达出更多内容特征，从而便于对其进行识别、检索与挖掘。从组织方式来看，相比于信息而言，知识更加结构化，也更具利用价值。我国《新一代人工智能发展规划》也提到要重点突破知识加工、深度搜索和可视交互核心技术 [80]，意味着基于知识的分析、挖掘及知识图谱构建将成为计算机与人工智能领域未来的重点发展方向。社交媒体与情感分析也是重要趋势之一，社交媒体上用户生成的海量文本信息与行为数据，为情感分析与用户行为研究提供了必要的数据基础，使得相关研究在计算机与人工智能技术的应用下得到了飞速发展。

表 1 2000-2020 年 *IPM* 突现词、突现强度及起止年份表

突现词	突现强度	起始年度	终止年度	突现时间示意图
information retrieval	12.05	2000	2009	
world wide Web	6.64	2000	2006	
document retrieval	4.17	2000	2004	
retrieval	3.44	2000	2005	
relevance	6.62	2001	2008	
interface	3.92	2002	2005	
design	3.32	2002	2010	
science	3.65	2003	2007	
information	3.46	2003	2005	
Internet	4.95	2004	2008	
cocitation	3.22	2004	2010	
query	4.44	2005	2009	
database	3.57	2005	2008	
user	5.37	2006	2011	
pattern	3.99	2008	2012	
search	3.42	2010	2013	
ranking	3.35	2012	2020	
graph	3.63	2016	2020	
impact	4.95	2017	2020	
knowledge	4.48	2017	2020	
Twitter	7.3	2018	2020	
sentiment analysis	6.14	2018	2020	
word of mouth	4.21	2018	2020	
emotion	3.44	2018	2020	

从期刊专辑也可以看出 *IPM* 关注的重点主题及其变化。在 2010 年以前，专辑主题基本围绕信息检索展开，例如在 2000 年的“基于网络的信息检索研究”（Web-based information retrieval research）专辑中，期刊编辑提到，互联网的发展扩大了信息检索的研究范围，研究人员开始逐渐关注网络信息检索以及信息检索系统的交互 [81]。同时，在 2000-2004 年间，国际计算机学会信息检索领域会议（ACM Special Interest Group on Information Retrieval，ACM SIGIR）连续举办了五届信息检索中的数学 / 形式化方法研讨会，证明了在信息检索中使用数学和形式化方法的重要性，*IPM* 也因此选编了相关论文，形成“信息检索中的数学模型设计、公式化和解释”（Model design, formulation and explanation in information retrieval using mathematics）专辑。步入新世纪后，互联网的迅猛发展、信息技术的更新迭代，以及数学统计方法的引进，为信息检索研究提供了源源不断的新动力，推动了研究内容向更加深入化发展，研究方法向更加技术化转变。

互联网的迅猛发展还带来了爆炸式增长的信息资源，而繁杂的文档信息中则可能包含许多具有重要价值的潜在知识，在传统的文本处理技术与工具不能满足新的用户需求的情况下，基于人工智能的文本挖掘方法应运而生，能够对浩瀚的文本资源进行有效的挖掘与利用 [82]。同时，随着 Facebook 与 Twitter 分别于 2004 年与 2006 年成立，社交媒体一跃成为便捷的交流工具和强大的自媒体平台，用户生成内容也因此成为网络信息资源的主要产生方式。由用户生成的信息资源难免掺杂许多个人的观点和情绪，这些主观性的言论大多会涉及社会热点事件或对产品 / 服务的消费评价，对此，情感分析迅速发展起来，并在舆情监测与商业营销等领域得到了广泛应用。针对文本挖掘和情感分析，*IPM* 也制作了相关的专辑，如“管理和挖掘多语言文档”（Managing and mining multilingual documents）、“社交和表达媒体中的情绪和情感”（Emotion and sentiment in social and expressive media）、“文本中的叙事提取”（Narrative extraction from texts）、“从社交网络中挖掘有价值的情报”（Mining actionable insights from social networks）等。总的来说，*IPM* 各个研究主题之间是互相联系的，主题的演进也与网络技术的发展、社交媒体的兴起有着密不可分的关系，同时，人工智能等新兴技术的出现也为计算机与信息科学领域带来了新的机遇。

5 结语

笔者通过词云图与共现网络图的绘制，将 *IPM* 近 20 年的研究主题划分为信息检索、文本分析、用户研究三大类。在信息检索方面，相关研究从信息检索模型的构建到搜索引擎与排名算法，全方位地讨论了信息检索的理论与方法，同时推进了图像检索技术的语义化发展。在文本分析方面，文本挖掘是主要的研究方向，在此基础上，社交媒体中的情感分析成为近期的研究热点，以知识图谱为依托的知识研究与分析也得到了持续的发展和应用。在用户研究方面，新型冠状病毒肺炎疫情发生后，健康信息搜寻、谣言识别与传播的相关研究受到更多关注，服务系统的个性化与人机交互的研究则凸显了以用户为中心的信息服务理念。

IPM 的主题演进主要呈现出 3 种特征：①始终以信息检索为核心主题。信息检索及以信息检索为基础的内容检索始终是 *IPM* 重点关注的主题。②从文本与信息分析向多媒体与知识分析转变。一方面，研究对象从文本信息向包含更多内容的多媒体信息拓展；另一方面，人工智能等新兴技术的发展，推动了信息分析向知识分析升级。③对用户情感的深入分析与挖掘。社交媒体上的用户生成内容催生了用户情感分析，使得用户研究向更深层次发展。

从主题分析结果来看，*IPM* 刊载论文关注的是计算机与信息科学领域的重点问题，使用的也是前沿的计算机技术与数学统计方法，能够从侧面展现出该领域在国际上的学术研究与

DOI: 10.13266/j.issn.2095-5472.2022.003

实践现状。然而，以单个期刊来反映整个学科领域的进展仍然具有局限性，后续研究可以对更多高影响力期刊进行计量研究与主题分析，以更全面地把握该学科领域的演化规律。

参考文献：

[1] 初景利 . 高端交流平台建设需要创新学术交流模式 [J]. 智库理论与实践 , 2021, 6(1): 7-9.

[2] DEHART F E. Monographic references and information science journal literature[J]. Information processing & management, 1992, 28(5): 629-635.

[3] TSAY M Y. A bibliometric analysis and comparison on three information science journals: JASIST, IPM, JOD, 1998–2008[J]. Scientometrics, 2011, 89(2): 591-606.

[4] 王曰芬 , 靳嘉林 . 比较分析《现代图书情报技术》近 10 年发文特征与发展趋势 [J]. 现代图书情报技术 , 2016(9): 1-16.

[5] 焦丽 . 我国信息检索研究综述 [J]. 情报探索 , 2007(6): 11-14.

[6] 王斌 . 信息检索导论 [M]. 北京 : 人民邮电出版社 , 2010.

[7] SPARCK JONES K, WALKER S, ROBERTSON S E. A probabilistic model of information retrieval: development and comparative experiments Part 1[J]. Information processing & management, 2000, 36(6): 779-808.

[8] SPARCK JONES K, WALKER S, ROBERTSON S E. A probabilistic model of information retrieval: development and comparative experiments Part 2[J]. Information processing & management, 2000, 36(6): 809-840.

[9] ROBERTSON S E, SPARCK JONES K. Relevance weighting of search terms[J]. Journal of the American Society for Information Science, 1976, 27(3): 129-146.

[10] 龚庆雄 . 基于实体的信息检索模型研究 [D]. 武汉 : 华中师范大学 , 2020.

[11] XU Z B, AKELLA R. Improving probabilistic information retrieval by modeling burstiness of words[J]. Information processing & management, 2010, 46(2): 143-158.

[12] DAHAK F, BOUGHANEM M, BALLA A. A probabilistic model to exploit user expectations in XML information retrieval[J]. Information processing & management, 2017, 53(1): 87-105.

[13] TAI X Y, REN F, KITA K. An information retrieval model based on vector space method by supervised learning[J]. Information processing & management, 2002, 38(6): 749-764.

[14] GUO J F, FAN Y X, PANG L, et al. A Deep Look into neural ranking models for information retrieval[J]. Information processing & management, 2020, 57(6): 102067.

[15] VAUGHAN L. New measurements for search engine evaluation proposed and tested[J]. Information processing & management, 2004, 40(4): 677-691.

[16] CAN F, NURAY R, SEVDIK A B. Automatic performance evaluation of Web search engines[J]. Information processing & management, 2004, 40(3): 495-514.

[17] KIM H, SEO J Y. High-performance FAQ retrieval using an automatic clustering method of query logs[J]. Information processing & management, 2006, 42(3): 650-661.

[18] JUNG S, HERLOCKER J L, WEBSTER J. Click data as implicit relevance feedback in Web search[J]. Information processing & management, 2007, 43(3): 791-807.

[19] WANG J M, PAN M, HE T T, et al. A Pseudo-relevance feedback framework combining relevance matching and semantic matching for information retrieval[J]. Information processing & management, 2020, 57(6): 102342.

[20] BIDOKI A M Z, YAZDANI N. DistanceRank: an intelligent ranking algorithm for Web pages[J]. Information processing & management, 2008, 44(2): 877-892.

[21] 孙君顶 , 原芳 . 基于内容的图像检索技术 [J]. 计算机系统应用 , 2011, 20(8): 240-244.

[22] HUANG P W, DAI S K. Design of a two-stage content-based image retrieval system using texture similarity[J]. Information processing & management, 2004, 40(1): 81-96.

[23] LU T C, CHANG C C. Color image retrieval technique based on color features and image bitmap[J]. Information processing & management, 2007, 43(2): 461-472.

[24] 图像检索：基于内容的图像检索技术 [EB/OL].[2021-09-17].https: //www.cnblogs.com/king-lps/p/11407206.html.

[25] PANDEY S, KHANNA P, YOKOTA H. A semantics and image retrieval system for hierarchical image databases[J]. Information processing & management, 2016, 52(4): 571-591.

[26] 袁军鹏 , 朱东华 , 李毅 , 等 . 文本挖掘技术研究进展 [J]. 计算机应用研究 , 2006(2): 1-4.

[27] 肖建国 . 试论文本挖掘及其应用 [J]. 图书馆学研究 ,

2008(4): 22-24.

[28] ALTINEL B, GANIZ M C. Semantic text classification: a survey of past and recent advances[J]. Information processing & management, 2018, 54(6): 1129-1153.

[29] ELNAGAR A, AL-DEBSI R, EINEA O. Arabic text classification using deep learning models[J]. Information processing & management, 2020, 57(1): 102121.

[30] MOHASSEB A, BADER-EL-DEN M, COCEA M. Question categorization and classification using grammar based approach[J]. Information processing & management, 2018, 54(6): 1228-1243.

[31] KASTRATI Z, IMRAN A S, YAYILGAN S Y. The impact of deep learning on document classification using semantically rich representations[J]. Information processing & management, 2019, 56(5): 1618-1632.

[32] HU G B, ZHOU S G, GUAN J H, et al. Towards effective document clustering: a constrained K-means based approach[J]. Information processing & management, 2008, 44(4): 1397-1409.

[33] CHEN C L, TSENG F S C, LIANG T. Mining fuzzy frequent item sets for hierarchical document clustering[J]. Information processing & management, 2010, 46(2): 193-211.

[34] ZHU S F, TAKIGAWA I, ZENG J. Field independent probabilistic model for clustering multi-field documents[J]. Information processing & management, 2009, 45(5): 555-570.

[35] FERSINI E, MESSINA E, ARCHETTI F. A probabilistic relational approach for Web document clustering[J]. Information processing & management, 2010, 46(2): 117-130.

[36] SOULIER L, TAMINE L, SHAH C. MineRank: leveraging users' latent roles for unsupervised collaborative information retrieval[J]. Information processing & management, 2016, 52(6): 1122-1141.

[37] KUCUKYILMAZ T, CAMBAZOGLU B B, AYKANAT C, et al. Chat mining: predicting user and message attributes in computer-mediated communication[J]. Information processing & management, 2008, 44(4): 1448-1466.

[38] TSENG Y H, LIN C J, LIN Y I. Text mining techniques for patent analysis[J]. Information processing & management, 2007, 43(5): 1216-1247.

[39] PONS-PORRATA A, BERLANGA-LLAVORI R, RUIZ-SHULCLOPER J. Topic discovery based on text mining techniques[J]. Information processing & management, 2007, 43(3): 752-768.

[40] 洪江涛 , 陈榴寅 , 黄沛 . 第三方点评网站对餐饮企业品牌形象与消费者行为的影响研究 —— 以大众点评网为例 [J]. 财贸经济 , 2013(10): 108-117.

[41] LIU B. Sentiment analysis and opinion mining[M]. San Rafael: Morgan & Claypool Publishers, 2012.

[42] 马力 , 宫玉龙 . 文本情感分析研究综述 [J]. 电子科技 , 2014, 27(11): 180-184.

[43] MOHAMMAD S M, ZHU X D, KIRITCHENKO S, et al. Sentiment, emotion, purpose, and style in electoral tweets[J]. Information processing & management, 2015, 51(4): 480-499.

[44] BALAHUR A, PEREA-ORTEGA J M. Sentiment analysis system adaptation for multilingual processing: the case of tweets[J]. Information processing & management, 2015, 51(4): 547-556.

[45] SEVERYN A, MOSCHITTI A, URYUPINA O, et al. Multi-lingual opinion mining on YouTube[J]. Information processing & management, 2016, 52(1): 46-60.

[46] AL-SMADI M, AL-AYYOUB M, JARARWEH Y, et al. Enhancing aspect-based sentiment analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features[J]. Information processing & management, 2019, 56(2): 308-319.

[47] KUMAR A, SRINIVASAN K, CHENG W H. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data[J]. Information processing & management, 2020, 57(1): 102141.

[48] MAHMOOD Z, SAFDER I, NAWAB R M A, et al. Deep sentiments in Roman Urdu text using recurrent convolutional neural network model[J]. Information processing & management, 2020, 57(4): 102233.

[49] 漆桂林 , 高桓 , 吴天星 . 知识图谱研究进展 [J]. 情报工程 , 2017, 3(1): 4-25.

[50] CHO H C, OKAZAKI N, MIWA M, et al. Named entity recognition with multiple segment representations[J]. Information processing & management, 2013, 49(4): 954-965.

[51] DERZYNSKI L, MAYNARD D, RIZZO G, et al. Analysis of named entity recognition and linking for tweets[J]. Information processing & management, 2015, 51(2): 32-49.

[52] TANG X, CHEN L, CUI J, et al. Knowledge

DOI: 10.13266/j.issn.2095-5472.2022.003

representation learning with entity descriptions, hierarchical types, and textual relations[J]. Information processing & management, 2019, 56(3): 809-822.
[53] BOUNHAS I, SOUDANI N, SLIMANI Y. Building a morpho-semantic knowledge graph for Arabic information retrieval[J]. Information processing & management, 2020, 57(6): 102124.
[54] LI D F, MADDEN A. Cascade embedding model for knowledge graph inference and retrieval[J]. Information processing & management, 2019, 56(6): 102093.
[55] SHIN S, JIN X, JUNG J, et al. Predicate constraints based question answering over knowledge graph[J]. Information processing & management, 2019, 56(3): 445-462.
[56] JANSSENS F, LETA J, GLANZEL W, et al. Towards mapping library and information science[J]. Information processing & management, 2006, 42(6): 1614-1642.
[57] 张一涵，袁勤俭．我国用户信息行为研究进展 [J]. 国家图书馆学刊，2014, 23(6): 91-98.
[58] 靳荣林．大学生自主学习情境下的信息搜寻行为影响因素探究 [D]. 保定：河北大学，2019.
[59] MAKRI S, BLANDFORD A, COX A L. Investigating the information-seeking behaviour of academic lawyers: from Ellis's model to design[J]. Information processing & management, 2008, 44(2): 613-634.
[60] JAMALI H R, NICHOLAS D. Interdisciplinarity and the information-seeking behavior of scientists[J]. Information processing & management, 2010, 46(2): 233-243.
[61] LYKKE M, PRICE S, DELCAMBRE L. How doctors search: a study of query behaviour and the impact on search results[J]. Information processing & management, 2012, 48(6): 1151-1170.
[62] PIAN W J, SONG S J, ZHANG Y. Consumer health information needs: a systematic review of measures[J]. Information processing & management, 2020, 57(2): 102077.
[63] ZHANG X F, GUO F, XU T X, et al. What motivates physicians to share free health information on online health platforms?[J]. Information processing & management, 2020, 57(2): 102166.
[64] HUVILA I, ENWALD H, ERIKSSON-BACKA K, et al. Anticipating ageing: older adults reading their medical records[J]. Information processing & management, 2018, 54(3): 394-407.
[65] JOHNSON J D. Health-related information seeking: is it worth it?[J]. Information processing & management, 2014, 50(5): 708-717.
[66] 赵宇翔，范哲，朱庆华．用户生成内容 (UGC) 概念解析及研究进展 [J]. 中国图书馆学报，2012, 38(5): 68-81.
[67] GE Y D, QIU J N, LIU Z Y, et al. Beyond negative and positive: exploring the effects of emotions in social media during the stock market crash[J]. Information processing & management, 2020, 57(4): 102218.
[68] LI L F, WANG Z Q, ZHANG Q P, et al. Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters[J]. Information processing & management, 2020, 57(6): 102313.
[69] LIU Y H, JIN X L, SHEN H W. Towards early identification of online rumors based on long short-term memory networks[J]. Information processing & management, 2019, 56(4): 1457-1467.
[70] ALKHODAIR S A, DING S H H, FUNG B C M, et al. Detecting breaking news rumors of emerging topics in social media[J]. Information processing & management, 2020, 57(2): 102018.
[71] WANG J, CLEMENTS M, YANG J, et al. Personalization of tagging systems[J]. Information processing & management, 2010, 46(1): 58-70.
[72] BELEM F M, MARTINS E F, ALMEIDA J M, et al. Personalized and object-centered tag recommendation methods for Web 2.0 applications[J]. Information processing & management, 2014, 50(4): 524-553.
[73] RENJITH S, SREEKUMAR A, JATHAVEDAN M. An extensive study on the evolution of context-aware personalized travel recommender systems[J]. Information processing & management, 2020, 57(1): 102078.
[74] WANG P L, HAWK W B, TENOPIR C. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach[J]. Information processing & management, 2000, 36(2): 229-251.
[75] KUMARAN G, ALLAN J. Adapting information retrieval systems to user queries[J]. Information processing & management, 2008, 44(6): 1838-1862.
[76] LOPATOVSKA I, ARAPAKIS I. Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction[J]. Information processing & management, 2011, 47(4): 575-592.
[77] COLE M J, GWIZDKA J, LIU C, et al. Inferring user knowledge level from eye movement patterns[J]. Information processing & management, 2013, 49(5):

1075-1091.
[78] CLARK M, RUTHVEN I, HOLT P O, et al. You have e-mail, what happens next? tracking the eyes for genre[J]. Information processing & management, 2014, 50(1): 175-198.
[79] HILBERINK-SCHULPEN B, NEDERSTIGT U, VAN MEURS F, et al. Does the use of a foreign language influence attention and genre-specific viewing patterns for job advertisements? an eye-tracking study[J]. Information processing & management, 2016, 52(6): 1018-1030.
[80] 中华人民共和国国务院 . 国务院关于印发新一代人工智能发展规划的通知 [EB/OL].[2021-10-26]. http: //www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
[81] SPINK A, QIN J. Introduction to the special issue on Web-based information retrieval research[J]. Information processing & management, 2000, 36(2): 205-206.
[82] 郭飞 . 文本挖掘方法探讨及应用 [D]. 成都 : 成都理工大学 , 2006.

作者贡献说明：
李涵霄：数据调研，撰写论文初稿；
杜杏叶：提出选题，修改论文及定稿。

Research Progress in Computer and Information Science in Recent 20 Years: Thematic Analysis of *Information Processing and Management*

Li Hanxiao[1,2] Du Xingye[1,2]
[1]National Science Library, Chinese Academy of Sciences, Beijing 100190
[2]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] This paper analyzes the themes of papers published in *Information Processing and Management* from 2000 to 2020, in order to know the thematic focus and evolution trends of *IPM*, and provide references for the development and related research of computer and information science. [**Method/process**] Firstly, based on 1852 research papers in the full-text database of ScienceDirect, this paper counted and visualized the titles, abstracts and keywords of the papers to classify the thematic categories. Then, this paper analyzed the research themes of each category systematically. At last, it compared the thematic focus in different periods and analyzed thematic evolution trends. [**Result/conclusion**] *IPM* mainly focuses on three themes of information retrieval, text analysis and user research, and presented evolution characteristics in general: information retrieval as the core theme, transform from text and information analysis to multimedia and knowledge analysis, and in-depth analysis and mining of user sentiments.

Keywords: *Information Processing and Management* *IPM* computer and information science thematic analysis